

局部修复码的最优构造

何亚锦¹, 孙伟¹, 沈克勤¹, 张鑫楠¹, 刘向阳²

(1. 长安大学 信息工程学院, 陕西 西安 710064;

2. 国防科技大学 信息通信学院, 陕西 西安 710106)

摘要: 分布式存储系统采用冗余策略来确保数据的可靠性和可用性, 局部修复码 (locally repairable codes, LRC) 引起了广泛的关注, 极大地减少了数据修复过程中所连接的节点数, 在数据存储中作用极大。每个信息码元可以从其他 t 个不相交的集合中修复, 且每个集合大小为 r , 称此类码具有 (r, t) 局部度。从校验矩阵入手, 提出两种构造具有 (r, t) 局部度的 LRC 的方法。方法一利用 $\lambda = 1$ 的非循环相对差集 (relative difference sets, RDS) 构造关联矩阵, 方法二提出了利用酉设计构造关联矩阵, 均在关联矩阵的右侧添加单位矩阵, 构造 LRC 的校验矩阵。两种方法构造的 LRC 均是一个修复集中包含一个校验节点, 并且可以达到任意 (r, t) 局部度。理论分析表明, 构造的两种码的最小距离均满足最小距离界, 证明了两种码均是最优的 LRC。非循环相对差集构造的码的信息率为 $1/2$, 酉设计构造的码的码率在一定条件下高于 $1/2$, 码率为 $\frac{r}{r+t}$ 。

关键词: 分布式存储系统; 局部修复码; 相对差集; 酉设计; 最小距离

中图分类号: TP301

文献标识码: A

文章编号: 1673-629X(2021)04-0112-06

doi: 10.3969/j.issn.1673-629X.2021.04.019

Optimal Construction of Locally Repairable Codes

HE Ya-jin¹, SUN Wei¹, SHEN Ke-qin¹, ZHANG Xin-nan¹, LIU Xiang-yang²

(1. School of Information Engineering, Chang'an University, Xi'an 710064, China;

2. School of Information and Communication, National University of Defense Technology, Xi'an 710106, China)

Abstract: The distributed storage system adopts redundancy strategy to ensure the reliability and availability of data. Locally repairable codes (LRC) has attracted much attention, which greatly reduce the number of nodes connected in the process of data repair and play a great role in data storage. Each information symbol can be repaired from t other disjoint sets and each set size is r , we call codes have (r, t) local degree. Starting with the calibration matrix, two methods of constructing LRC with (r, t) local degree are proposed. The first one uses acyclic relative difference sets (RDS) with parameters $\lambda = 1$ to construct the correlation matrix, and the other one proposes to construct the correlation matrix by using unitary design, then adding the unit matrix to the right of the correlation matrix, and constructing the LRC check matrix. The LRC constructed by both methods are that a repair group contains a check node and can achieve arbitrary (r, t) local degree. Theoretical analysis shows that the minimum distance of the constructed two codes satisfies the minimum distance boundary, which proves that both codes are optimal LRC. The information rate of code constructed by acyclic relative difference sets is $1/2$, and that of code constructed by unitary design is higher than $1/2$ under certain conditions, and the code rate is $\frac{r}{r+t}$.

Key words: distributed storage system; locally repairable codes; relative difference sets; unitary design; minimum distance

0 引言

随着计算机存储能力的提升, 近年来数据呈指数型增长, 如何可靠存储这些数据引起人们的思考。现如今, 社交媒体服务和云服务的用户经常上传图像和视频等大型数据文件, 需要巨大的存储空间, 这种存储

空间是以分布式存储的形式实现的^[1-2]。分布式存储系统存储规模大, 节点数量增长迅速, 无法避免磁盘故障和节点失效等情况。为了保证分布式存储系统在节点故障时的正常工作, 复制策略和纠删码策略被提出并在实际系统中应用^[3-4]。

收稿日期: 2020-06-06

修回日期: 2020-10-10

基金项目: 陕西省自然科学基金资助项目 (2019JM-386); 中央高校基本科研业务费专项资金资助项目 (300102248104, 300102248201, 300102248401); 大学生创新创业训练计划资助项目 (201910710071)

作者简介: 何亚锦 (1996-), 女, 硕士, 研究方向为分布式存储和网络编码。

复制策略较为简单,通过对数据进行直接复制得到多个副本,最常见的是三副本复制。由于复制策略是对所有的数据复制若干倍,增加了系统的存储开销。当前数据海量化增长,复制策略使得存储开销成本变得更高。纠删码策略是将原文件分成 k 个等大小的数据块,编码生成 n 个数据块,分别存储在 n 个不同的节点上。纠删码策略解决了存储开销大这一缺点,保证了数据传输的可靠性,但是也有一定的弊端。考虑故障节点的修复开销,传统复制策略中,只需要复制数据到新的存储节点上就可以修复故障节点了,修复过程简单。而纠删码策略在对故障节点进行修复时,需要对 k 个数据块重新编码生成原文件再进行编码,这个过程涉及的计算复杂度高,修复带宽开销明显比传统的复制策略高。

为了减小节点故障时的修复开销,再生码(regenerating codes)在文献[5]中被提出,但在修复过程中需要连接的节点数目较多,计算复杂。为了减小所涉及的节点数目,即磁盘 I/O 开销,文献[6-7]提出了局部修复码(locally repairable codes, LRC),只需要连接少量的节点就能完成故障节点的修复,明显降低了修复成本,提高了修复效率。如果一个码元最多可以被 r 个其他码元来修复,就称该码具有局部度 r 。局部修复码因其高效的修复效率和低的修复成本,在分布式系统中广泛应用。在分布式文件 Hadoop 中,文献[8]中的 Face book 和文献[9]中的 Windows Azure 存储使用的都是 LRC。文献[10]采用多项式运算的方法构造 LRC,使得节点修复时仅需要连接 2 或 3 个节点。文献[11]构造的 LRC 采用双层编码结构,编码复杂且节点修复复杂度较高。

许多学者开始对二元 LRC 的构造展开研究。文献[12]提出了 (r, t) 局部度的概念,即对于局部修复码来说,一个故障节点可以被其他 t 个不相集的节点分别修复,并且修复集的大小为 r ,则称 (r, t) 局部修复码。文献[13]提出了基于最小距离 $d = 4$ 的 (n, k, d, r) LRC 码的构造,但其需满足 $r + 1$ 整除 n 的条件。文献[14]构造的 LRC 的最小距离 $d \geq 6$,局部性 $r \geq 2$ 。文献[15]中构造的 LRC 码的局部性为 $r = 2$ 和 $r = 3$ 。

针对上面构造方法的复杂性和参数限制较大的问题,该文提出构造新二元 LRC 的两种方法,方法一利用非循环相对差构造一个二元正则校验矩阵,方法二利用西设计定义一个关联矩阵。

实验表明,基于非循环相对差集构造的码的码率为 $1/2$;基于西设计构造的码率为 $\frac{r}{r+t}$,且在一定条件下,码率高于 $1/2$ 。

1 相关概念

在本节中,给出了一些关于 LRC 的基础知识,介绍了 LRC 参数的一定界限,描述了 LRC 具有给定参数的最优性。

1.1 系统码

大小为 M 的文件被划分为 k 个大小相等的块,编码成 n 个编码块,最小距离为 d ,记为 (n, k, d) 码。最大距离可分码(MDS)的最小距离满足 $d = n - k + 1$ 。一个 (n, k, d) MDS 码从 n 个编码块中的任取 k 个编码块来修复一个失效编码块。用 r 表示校验块的数目,其中 $r = n - k$ 。如果系统线性码 C 是一个 (n, k, d) 码,则生成矩阵具有以下形式: $G = [I_k, P] \in F_q^{k \times n}$,其中 $P \in F_q^{k \times (n-k)}$ 。进一步可以得到码 C 的校验矩阵 $H = [-P^T, I_{n-k}] \in F_q^{(n-k) \times n}$ 。

1.2 局部性修复码

设有限域 F_q 含有 q 个元素,令 $[n] = \{1, 2, \dots, n\}$,如果码 $c \in C$ 的每个编码都可以从码 c 的 r 个其他编码中恢复,则称码 $C \subset F_q^n$ 具有局部性 r 。换句话说,这意味着,给定 $c \in C$,存在相应的集合 $R_i \subset [n] \setminus i$, $|R_i| \leq r$ 的子集能够恢复 c_i ,其中 $i \in [n]$,则称子集 R_i 为 c_i 的修复集。

1.3 (r, t) 可用性

假设码 C 的每个编码都可以从大小为 r 的 t 个不相交子集中恢复,就称码 C 具有 (r, t) 局部度。在下面正式定义具有 (r, t) 局部度的线性码。

对于线性码 (n, k, d) ,第 i 个编码 $c_i (1 \leq i \leq n)$ 具有 (r, t) 局部度,满足以下条件:

- (1) 存在 t 个子集 $R_1^{(i)}, R_2^{(i)}, \dots, R_t^{(i)} \subseteq [n] \setminus i$,使得 c_i 可以被 $R_j^{(i)} (1 \leq j \leq t)$ 中的码元修复;
- (2) 对所有 $1 \leq j \leq t$, $|R_j^{(i)}| \leq r$;
- (3) 对任意 $j \neq l$, 且 $1 \leq j, l \leq t$, 有 $R_j^{(i)} \cap R_l^{(i)} = \emptyset$ 。

其中, $R_j^{(i)} (1 \leq j \leq t)$ 称为 c_i 的修复集合。如果仅是其信息位具有 (r, t) 局部性,并且一个校验码元仅在一个修复集中,则称该码具有 (r, t) 局部性,记为 (n, k, r, t) LRC。

1.4 最小距离

码的最小距离是两个码字之间的所有元素取不同时的元素个数的最小值,即 $d = \min\{d(u, v)\}$,其中 $d(u, v)$ 是两个码字 u, v 的汉明距离,对于线性码 (n, k, d) ,出现 d 个节点故障时就不能获取原始文件了。

对于单校验 (n, k, r, t) LRC,最小距离满足 $d \geq t + 1$,文献[16]给出了码的最小距离的上界:

$$d \leq n - k + 2 - \left\lceil \frac{t(k-1)+1}{t(r-1)+1} \right\rceil \quad (1)$$

文献[17]单校验 (n, k, r, t) LRC,给出了一个码

的最小距离的上界:

$$d \leq n - k - \lceil \frac{kt}{r} \rceil + t + 1 \quad (2)$$

2 最优局部修复码的构造

2.1 基于非循环相对差集的二元单校验 \$(r, t)\$ LRC

本节旨在构造 \$(r, t)\$ 正则矩阵, 首先给出了相对差集的定义, 又进一步给出 LRC 校验矩阵的具体构造方式。

定义 1(相对差集)^[18]: 设 \$R\$ 为一个 \$uv\$ 阶群, \$N\$ 是它的一个 \$u\$ 阶子群, 而 \$D\$ 为一个 \$R\$ 的 \$k\$ 子集。如果对于 \$R \setminus N\$ 的每一个元素 \$a\$, 都存在 \$\lambda\$ 对 \$D\$ 中的两个不同元素 \$(d_i, d_j)\$, 使得 \$a = d_i d_j^{-1}\$; 并且 \$N\$ 中的任意元素都不能表示为 \$d_i d_j^{-1}\$ 的形式, 其中 \$d_i, d_j \in D\$, 那么就称 \$D\$ 为一个 \$(v, u, k, \lambda)\$-相对差集 (relative difference set, RDS), 而 \$N\$ 是其相对应的禁子群。

在下面的描述中, 符号 \$\oplus\$ 表示直接求和, \$Z_p\$ 表示模 \$p\$ 的加群。让 \$G_n\$ 是 \$p^n\$ 阶阿贝尔群的元素, 即 \$Z_p\$ 的 \$n\$ 元组。RDS 的具体构造过程如下:

首先, 令 \$G = Z_p \oplus G_n\$, \$N = Z_p \oplus \{0\}\$, \$D = \{(f(m), m) \mid m = (m_1, m_2, \dots, m_n) \in G_n\}\$, 得到一个 \$G\$ 相对于 \$N\$ 的相对差集即 \$(p^n, p, p^n, p^{n-1})\$, \$f(m)\$ 如下:

$$f(m) = \sum_{i=1}^n a_i m_i^2 (\text{mod } p) \quad (3)$$

$$a_i \neq 0 (\text{mod } p), i = 1, 2, \dots, n$$

由于感兴趣的 \$\lambda = 1\$ 的相对差集, 因此令上述公式中 \$n = 1\$, 进一步得到 \$f(m) = a_1 m^2 (\text{mod } p)\$, \$m \in Z_p\$, 得到相对差集 \$D(p, p, p, 1)\$。

由于 \$D\$ 的元素之间没有明确的顺序, 在构造 LRC 时候假定一个字典顺序定义为:

$$i = f(m) \times p + m, m \in Z_p, i \in (0, p^2 - 1) \quad (4)$$

构造 1: 设非循环 RDS \$D(p, p, p, 1)\$, 其中 \$p\$ 是素数。对于 \$(i_1, i_2), (j_1, j_2) \in G\$, 得到字典顺序 \$i = p \times i_1 + i_2\$ 和 \$j = p \times j_1 + j_2\$, 则校验矩阵 \$\mathbf{H} = [I(C) \ I]\$。其中 \$I(C)\$ 定义如下: \$I(C) = [c_{i,j}]_{0 \leq i,j < p^2}\$

$$c_{i,j} = \begin{cases} 1 & (j_1, j_2) \in D + (i_1, i_2) \\ 0 & \text{其他} \end{cases}$$

$$D + (i_1, i_2) = \{(d + i_1, d + i_2) \mid (d_1, d_2) \in D\} \quad (5)$$

由校验矩阵 \$\mathbf{H} = [I(C) \ I]\$ 构造的局部修复码参数为 \$n = 2p^2, k = p^2, r = p, t = p\$。

定理 1: 基于非循环相对差集 \$D(p, p, p, 1)\$ 构造的局部修复码 \$(n = 2p^2, k = p^2, r = p, t = p)\$ 是最优的, 满足边界条件。

证明: 基于非循环相对差集构造的码的长度为 \$2p^2\$, 维数为 \$p^2\$, \$I\$ 的每一列对应着奇偶校验位, \$I(C)\$

的每一列代表信息位, \$I(C)\$ 的每一列码重为 \$p\$, 也就是说每个信息符号有 \$t = p\$ 个修复集。每一行的行重为 \$p\$, 并且任意两行之间最多有一个公共的位置, 因此局部性 \$r = p\$。每个修复集只包含一个校验位。注意校验矩阵的每一行的行重为 \$p + 1\$, 又因为 \$d \geq t + 1 = p + 1\$。因此最小距离 \$d = p + 1\$, 很明显构造出的码是最优的。进一步地, 由码率 \$R\$ 的计算公式可以得到 \$R = k/n = 1/2\$。

$$p + 1 = d \leq n - k - \lceil \frac{kt}{r} \rceil + t + 1 = p + 1 \quad (6)$$

例 1: 令 \$Z_3 = \{0, 1, 2\}\$, \$G = Z_3 \oplus Z_3, p = 3, n = 1, a_1 = 1, f(m) = m^2 (\text{mod } 3)\$, 计算 \$D = \{(f(m), m) \mid m \in \{0, 1, 2\}\}\$ 得到 \$D(3, 3, 3, 1)\$, 即 \$D = \{(0, 0), (1, 1), (1, 2)\}\$。对于 \$G = Z_3 \oplus Z_3\$ 中的所有元素, 有以下 \$D\$ 的集合:

$$\begin{aligned} D + (0, 0) &= \{(0, 0), (1, 1), (1, 2)\} \\ D + (0, 1) &= \{(0, 1), (1, 2), (1, 0)\} \\ D + (0, 2) &= \{(0, 2), (1, 0), (1, 1)\} \\ D + (1, 0) &= \{(1, 0), (2, 1), (2, 2)\} \\ D + (1, 1) &= \{(1, 1), (2, 2), (2, 0)\} \\ D + (1, 2) &= \{(1, 2), (2, 0), (2, 1)\} \\ D + (2, 0) &= \{(2, 0), (0, 1), (0, 0)\} \\ D + (2, 1) &= \{(2, 1), (0, 2), (0, 0)\} \\ D + (2, 2) &= \{(2, 2), (0, 0), (0, 1)\} \end{aligned}$$

通过使用字典顺序, 可以得到关联矩阵:

$$I(C) = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

由校验矩阵 \$\mathbf{H} = [I(C) \ I]\$ 定义的单校验 \$(18, 9, 3, 3)\$ LRC, 最小距离 \$d = t + 1 = 4\$。通过式 (2) 可知, 它是最优的 LRC, 且码率 \$R = k/n = 1/2\$。

2.2 基于西设计构造 \$(r, t)\$ LRC

西设计是基于射影几何构造的 Steiner 设计, 在本节首先介绍西设计的定义, 进一步利用西设计构造稀疏 \$(r, t)\$ 正则矩阵, 给出 LRC 校验矩阵的具体构造方式, 并证明得到的码是最优的。

考虑有限域 \$\text{GF}(q)\$ 的三元组 \$x = (x, y, z)\$ 的集合 \$S\$, 其中 \$(x, y, z)\$ 不全为零。\$S\$ 有 \$q^3 - 1\$ 个元素, 但是对于一些非零元素 \$\lambda \in \text{GF}(q)\$, 如果满足 \$x = \lambda y\$, 定义三元组 \$x\$ 和 \$y\$ 是等价的。用 \$[x]\$ 表示 \$x\$ 的等价类, 每个

等价类都有 $q-1$ 个成员,对应于 $q-1$ 个可能的非零值 λ , 因此有 $\frac{q^3-1}{q-1} = q^2+q+1$ 不同类 $[x]$, 将其作为 $PG(2, q)$ 的点。考虑三元组 $L = (l_1, l_2, l_3) \neq (0, 0, 0)$, 射影几何中的线 $[L]$ 是过点 $[p] = (p_1, p_2, p_3)$, 当且仅当 $p_1 l_1 + p_2 l_2 + p_3 l_3 = 0$ 。由上可知, 有 q^2+q+1 个块, 每条线 $[L]$ 上有 $q+1$ 个不同的点 $[p]$ 。

酉设计构造: 设 $P = PG(2, q)$ 用齐次坐标表示, 酉设计上的点 $(x:y:z)$ 满足 $xx^q + yy^q + zz^q = 0$, 称这个酉设计是一条 Hermitian 曲线。

例2: 要构造一个酉设计, 在偶数阶 $q = m^2$ 的射影平面 $PG(2, q)$ 有 m^3+1 个点, 每条线有 $m+1$ 个点。在 $PG(2, m^2)$ 中, Hermitian 曲线酉设计上的点 (x, y, z) 满足下面性质 $f(x, y, z) = x^{m+1} + y^{m+1} + z^{m+1} = 0$ 。由上面给出的平面 $PG(2, 2^2)$ 点的集合满足 $x^3 + y^3 + z^3 = 0$, 点集 $110, 011, 1\alpha 0, 01\alpha, 1\beta 0, 01\beta, 101, 10\alpha, 10\beta$ 线 $[11\beta], [010], [1\beta 1], [\beta\alpha 1], [1\alpha 1], [111], [001], [\beta 11], [100], [11\alpha], [1\alpha\beta]$ 。所有这些都包含酉极性中的3个点, 所有其他线都包含这些点之一。酉设计具有酉极性的点集作为点, 对于射影平面中满足酉极性的点集的线, 都包含酉极性中的 $m+1$ 个点。

构造2: 酉设计参数如下:

$$v = m^3 + 1, b = \frac{m^2(m^3 + 1)}{(m + 1)} = m^2(m^2 - m + 1), k = m + 1, r = m^2 \quad (7)$$

酉设计的 $v \times b$ 的关联矩阵如下所示, 其中每一行表示点 P_i , 每一列表示块 B_j :

$$I(C) = \begin{cases} 1 & P_i \in B_j \\ 0 & \text{其他} \end{cases} \quad (8)$$

由校验矩阵 $H = [I(C) \ I_r]$ 构造的 LRC 的参数如下所示: $n = b + v = m^4 + m^2 + 1, k = b = m^2(m^2 - m + 1), r = m^2, t = m + 1$ 。

定理2: 基于酉设计构造的局部修复码 $n = b + v = m^4 + m^2 + 1, k = b = m^2(m^2 - m + 1), r = m^2, t = m + 1$ 是最优的, 且满足边界条件是最优的。

证明: 酉设计构造的 $v \times b$ 的关联矩阵, 每行有 m^2 个1, 并且任意两行之间最多有一个公共的位置, 因此局部性 $r = m^2$; 每列有 $m+1$ 个1, 说明每个数据块有 $t = m+1$ 个修复集。在此关联矩阵右侧添加单位矩阵, 可知校验矩阵的每一行的行重为 m^2+1 , 又因为有 $d \geq t+1 = m+2$ 。因此最小距离 $d = m+2$, 很明显构造出的码是最优的。进一步地, 可以计算其码率 R 如下:

$$R = \frac{k}{n} = \frac{b}{b+v} = \frac{1}{1+\frac{v}{b}} = \frac{1}{1+\frac{m^3+1}{m^2(m^2-m+1)}} =$$

$$\frac{m^2}{m^2+m+1} = \frac{r}{r+t} \quad (9)$$

例3: 由例2 射影平面 $PG(2^2)$ 上的酉极性导出的酉设计 $(9, 12, 4, 3)$, 如下所示:

1. 相应的区组设计如下:

$[0, 1, 5], [0, 2, 4], [0, 3, 6], [0, 7, 8],$
 $[1, 2, 3], [1, 4, 7], [1, 6, 8], [2, 5, 8],$
 $[2, 6, 7], [3, 4, 8], [3, 5, 7], [4, 5, 6]$

2. 关联矩阵:

$$I(C) = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

由校验矩阵 $H = [I(C) \ I_r]$ 构造的单校验 $(21, 12, 4, 3)$ LRC, 最小距离 $d = t + 1 = 4$, 码率 $R = k/n = 4/7$ 。

3 性能分析

由于 LRC 主要考虑其最小距离和码率的问题, 文中提出的两种码很容易证明满足最小距离界, 并进一步证明是最优的, 所以本节主要从码率和故障节点修复两方面进行了比较。

3.1 码率

本节主要从码率方面与可达任意 (r, t) 局部度的 LRC 进行了比较。首先在表1中给出了文中构造的两类码与文献[19]中的直积码的构造参数, 并为了方便观察其大小, 给出了如图1所示的曲线。

表1 局部修复码构造参数

	n	k	d	R
构造1	$2r^2$	r^2	$t+1$	$1/2$
构造2	$m^4 + m^2 + 1$	$m^2(m^2 - m + 1)$	$t+1$	$r/(r+t)$
直积码	$(r+1)^t$	$(r+1)^t$	$t+1$	$[r/(r+1)]^t$

文献[19]中的直积码, 是由 t 个二元单校验 $(r+1, t)$ 码生成的 LRC, 码率为 $(\frac{r}{r+1})^t$ 。当 $t > 1$ 时,

$(1 + \frac{1}{r})^t > 1 + \frac{t}{r}$, 因此可得:

$$(\frac{r}{r+1})^t = \frac{1}{(1 + \frac{1}{r})^t} < \frac{1}{1 + \frac{t}{r}} = \frac{r}{r+t}$$

经上面计算分析可知, 局部度 r 相同时, 当 $t > 1$

时,基于西设计构造的码的码率比直积码的码率高。当 $r > t$ 时,基于西设计构造的码的码率高于 $1/2$ 。

为了更加直观地观察表中三种码的码率的大小,给出了 $r = 3$ 时, $t - k/n$ 的曲线,如图 1 所示。

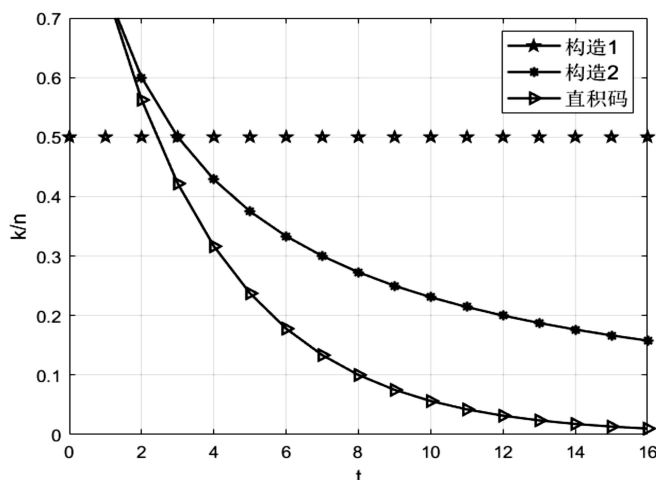


图 1 $r = 3$, 三种码的码率与可用性 t 的关系

由图 1 可知,当 $r = 3$,构造 2 的码率是恒大于直积码的码率,而构造 1 的码率在一定条件下也是高于直积码的码率。

3.2 故障节点修复

节点故障可分为单节点故障和多节点故障。考虑单个编码块故障的情况,又可以分为单个数据块故障和单个校验块故障。针对单个数据块故障的时候,可以选取任意一个修复集就可以完成修复,可选择性较多;针对校验块故障,只需要通过相关的数据块通过异或即可修复。单节点故障修复较为简单,接下来考虑多节点故障的情形。

对于两个或多个故障块,修复可以分为并行修复和顺序修复。并行修复是指两个故障块分开各自利用现有的完整的编码块进行修复;而顺序修复是指先修复好一个故障块,然后利用修复好的一个故障块和现有的编码块修复另一个故障块。由于每个修复集中只含有一个校验数据块,因此优先选取并行修复方案。如果不能选取并行修复方案,则选取顺序修复方案。考虑到一个修复集中全部节点故障发生的概率较小,因此并行修复方案为主要方案。

以例 1 说明,由校验矩阵构造的二元局部修复码的构造方法是将原始文件 M 平均分成 9 个大小相等的数据块,编码生成 18 个编码块,其中令 (c_1, c_2, \dots, c_9) 为数据块, $(c_{10}, c_{11}, \dots, c_{18})$ 为校验块。单数据块故障时,可以从校验矩阵中得到数据块 c_1 的修复关系,即 $c_1 = c_{10} - c_5 - c_6 = c_{17} - c_3 - c_8 = c_{18} - c_9 - c_2$ 。因此可以表示 c_1 的修复集 $R_1^{(1)} = \{5, 6, 10\}$, $R_2^{(1)} = \{3, 8, 17\}$, $R_3^{(1)} = \{2, 9, 18\}$, 同样地,也可以得到其他数据块的修复集即 $R_1^{(2)} = \{4, 6, 11\}$, $R_2^{(2)} = \{3, 7, 16\}$, $R_3^{(2)} = \{1, 9, 18\}$ 等等。如果校验块 c_{10} 故障,只需要将 $c_1, c_5,$

c_6 进行异或就能修复。可见单节点修复过程比较简单。

针对例 1 中的编码块,若出现多个编码块故障时,如果 c_1 和 c_{10} 故障,此时由于数据块和校验块同时故障且在一个修复集中,则这种情况只能选择顺序修复,即先在修复集 $R_2^{(1)} = \{3, 8, 17\}$ 或 $R_3^{(1)} = \{2, 9, 18\}$ 中任取一个进行修复故障块 c_1 ,然后利用 c_1, c_5, c_6 进行异或来修复 c_{10} 。如果数据块 c_1 和 c_2 同时故障,则可以优先选取并行修复的方案,即同时从 c_1 的修复集 $R_1^{(1)} = \{5, 6, 10\}$ 或 $R_2^{(1)} = \{3, 8, 17\}$ 中修复 c_1 ,从 c_2 的修复集中 $R_1^{(2)} = \{4, 6, 11\}$ 和 $R_2^{(2)} = \{3, 7, 16\}$ 任选一个来修复 c_2 。多节点故障时,局部修复码采用并行修复的修复方案,各编码块同时进行修复,减少了修复时间,并提高了修复效率。

4 结束语

由于局部修复码能够解决分布式存储系统中常见的问题,如修复多个节点故障和管理热数据,近年来受到了广泛的关注。构造适用于分布式存储系统的 LRC 成为研究的一个方向。局部修复码的研究主要从生成矩阵和校验矩阵入手,该文从校验矩阵的角度构造相应的二元最优单校验 LRC。首先利用相对差集和西设计构造了两种关联矩阵,再在此矩阵的右侧添加单位矩阵,来构造所需的校验矩阵。实验表明,构造的码均能达到最小距离界限,并且在每个修复集中只包含一个校验块,证明 LRC 是最优的。进一步给出了其码率大小,基于非循环差集构造的码率为 $1/2$,基于西设计的 LRC 的码率可达 $\frac{r}{r+t}$ 。

在分布式存储系统中,由于局部修复码在节点故障修复时所连接的节点数小于其维数 k ,实用性价值

高。对一些不经常使用的数据,使用局部修复码能节约很多修复成本。在对那些经常使用的数据时,采用二元局部修复码,编码过程相对简单,在修复时可进行并行修复,修复效率高。利用 (r, t) 局部修复码,可以容忍多节点故障并对其修复采用并行修复,且在修复过程中所连接的节点数大大减少,减少了修复复杂度和修复时间。可见所提出的二进制 LRC 在系统实现方面有很大的研究价值。该文只给出了各节点存储数量相同局部修复码的构造,构造节点存储容量不同的局部修复码将是接下来需要考虑的问题之一。

参考文献:

- [1] 余春雷,王 静,王 秘,等. 图因子分解的部分重复码构造[J]. 中国科技论文,2019,14(11):1260-1264.
- [2] 王甜甜,余春雷. 异构分布式存储系统中部分重复码的构造[J]. 信息通信,2019(1):111-112.
- [3] 宫 婧,王文君. 大数据存储中的容错关键技术综述[J]. 南京邮电大学学报:自然科学版,2014,34(4):20-25.
- [4] 曾赛峰,屈喜龙. 云存储环境下分组校验纠删码冗余算法研究[J]. 湖南工程学院学报:自然科学版,2016,26(4):42-45.
- [5] DIMAKIS A G, GODFREY P B, WU Y, et al. Network coding for distributed storage systems[J]. IEEE Transactions on Information Theory, 2010, 56(9):4539-4551.
- [6] 刘振元. 基于纠删码的分布式存储访问优化技术研究[D]. 长沙:国防科学技术大学,2015.
- [7] 张志芳,徐敬可,刘木兰. 构造小域上的最优局部修复码[J]. 中国科学:数学,2017,47(11):1607-1614.
- [8] HUANG C, SIMITCI H, XU Y, et al. Erasure coding in windows azure storage [C]//Proceedings of the 2012 USENIX conference on annual technical conference. Boston, MA: USENIX Association, Berkeley, 2012:2.
- [9] SATHIAMOORTHY M, ASTERIS M, PAPAILIOPOULOS D, et al. XORing elephants: novel erasure codes for big data [J]. Proceedings of VLDB Endowment, 2013, 6(5):325-336.
- [10] RAWAT A S, VISHWANATH S. On locality in distributed storage systems[J]. Mathematics, 2012, 15(2):497-501.
- [11] XIE H, YAN Z. Two-layer locally repairable codes for distributed storage systems [C]//2014 IEEE international conference on communications (ICC). Sydney, NSW: IEEE, 2014: 3902-3907.
- [12] WANG A, ZHANG Z. Repair locality with multiple erasure tolerance[J]. IEEE Transactions on Information Theory, 2014, 60(11):6979-6987.
- [13] SHAHABINEJAD M, KHABBAZIAN M, ARDAKANI M. A class of binary locally repairable codes[J]. IEEE Transactions on Communications, 2016, 64(8):3182-3193.
- [14] NAM M Y, SONG H Y. Binary locally repairable codes with minimum distance at least six based on partial t-spreads[J]. IEEE Communications Letters, 2017, 21(8):1683-1686.
- [15] SILBERSTEIN N, ZEH A. Optimal binary locally repairable codes via anticodes [C]//2015 IEEE international symposium on information theory (ISIT). Hong Kong: IEEE, 2015:1247-1251.
- [16] RAWAT A S, PAPAILIOPOULOS D S, DIMAKIS A G, et al. Locality and availability in distributed storage [J]. IEEE Transactions on Information Theory, 2016, 62(8):4481-4493.
- [17] HAO J, XIA S T. Constructions of optimal binary locally repairable codes with multiple repair groups[J]. IEEE Communications Letters, 2016, 20(6):1060-1063.
- [18] 潘红艳,付绍静,杜 蛟. 一类基于相对差集的异步跳频序列系统[J]. 河南师范大学学报:自然科学版,2016,44(6):170-174.
- [19] CHAICHANAVONG P, SIEGEL P H. Relaxation bounds on the minimum pseudo-weight of linear block codes [C]//International symposium on information theory. Adelaide, SA: IEEE, 2005:805-809.